χ^2 Test for TV regularization parameter selection

Jodi Mead
Department of Mathematics
Boise State University

This work is supported by NSF-DMS-1043107

Inverse Problem

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\epsilon}$$

$$\mathbf{y} \in \mathbb{R}^m,\ \mathbf{A} \in \mathbb{R}^{m\times n},\ \mathbf{x} \in \mathbb{R}^n,\ \mathbb{E}(\boldsymbol{\epsilon}) = \mathbf{0},\ \mathrm{cov}(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}$$

Total Variation minimization

$$\hat{\mathbf{x}}_{tv} \in \arg\min_{\mathbf{x}} \frac{\lambda}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \|\nabla\mathbf{x}\|_1$$

Choice of λ example

$\lambda = 500$

$\lambda = 50000$

$\lambda = 500000$

TV regularization parameter selection

Approaches

- **L-curve**: no guarantee that norm of the data residual vs. the solution will be L-shaped.
- TV function viewed or approximated with a quadratic functional: **Discrepancy principle**[*], **Unbiased Predictive Risk Estimator (UPRE)**[**], **Generalized Cross Validation (GCV)**[***].
- **Stein's unbiased risk estimate (SURE)**[†] that relies on degrees of freedom estimates in the predictive risk estimator.

---

[*] Wen et. al, 2012; [**] Lin et. al, 2012; [***] Liao et. al, 2009
[†] Deledalle et. al, 2014

Residual properties

$$\mathbb{E}(\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2) = m\sigma^2$$

suggests solving nonlinear equation

$$f(\lambda) = \|\mathbf{y} - \mathbf{A}\mathbf{x}(\lambda)\|_2^2 - m\sigma^2 = 0$$

for λ. However

$$\hat{f}(\lambda) = \|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}(\lambda)\|_2^2 - m\sigma^2$$

is biased so choosing λ by solving $\hat{f}(\lambda) = 0$ leads to oversmoothing (Discrepancy principle).

Effective Degrees of Freedom (EDF)

Tikhonov regularization

$$\hat{\mathbf{x}}_{ls} \in \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \alpha^2 \|\mathbf{D}\mathbf{x}\|_2^2$$

gives ridge regression estimator

$$\hat{\mathbf{x}}_{ls} = (\mathbf{A}^T\mathbf{A} + \alpha^2 \mathbf{D}^T\mathbf{D})^{-1}\mathbf{A}^T\mathbf{y}.$$

Predictors are

$$\mathbf{A}\hat{\mathbf{x}}_{ls} = \mathbf{N}(\alpha)\mathbf{y}$$

and*

$$\mathbb{E}(\|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}_{ls}\|_2^2) = (m - \mathrm{tr}\mathbf{N}(\alpha))\sigma^2$$

---

*Hall et. al, 1987.

Degrees of Freedom in Nonlinear (TV) Regression

Tikhonov regularization term defines smoothing matrix $\mathbf{A}\hat{\mathbf{x}}_{ls} = \mathbf{N}\mathbf{y}$ while nonlinear smoothers (e.g. TV) have

$$\mathbf{A}\hat{\mathbf{x}}_{tv} = \delta(\mathbf{y}).$$

Degrees of freedom of δ are given by*

$$df(\mathbf{A}\hat{\mathbf{x}}_{tv}) = \sum_{i=1}^{m} \text{cov}((\hat{\mathbf{x}}_{tv})_i, \mathbf{y}_i)/\sigma^2$$

e. g. $\mathrm{df}(\mathbf{A}\hat{\mathbf{x}}_{ls}) = \mathrm{tr}(\mathbf{N})$.

---

*Efron, 2004.

Degrees of Freedom for TV (generalized Lasso)[*,**]

$$\hat{\mathbf{x}}_{tv} \in \arg\min_{\mathbf{x}} \frac{\lambda}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \|\mathbf{D}\mathbf{x}\|_1$$

No assumptions on $\mathbf{A}$ or $\mathbf{D}$:

$$df(\mathbf{A}\hat{\mathbf{x}}_{tv}) = \mathbb{E}[dim(\mathbf{A}(null(\mathbf{D}_{-\mathcal{A}})))], \quad \mathcal{A} = \{i : (\mathbf{D}\hat{\mathbf{x}}_{tv})_i \neq 0\}$$

If $\mathbf{A}$ is full rank, the number of non-zero predictors $\mathbf{D}\hat{\mathbf{x}}_{tv}$ is an unbaised estimate for $df(\mathbf{A}\hat{\mathbf{x}}_{tv})$.

[*] Tibshirani 2012;[**] Dossal 2013.

χ^2 Functional for TV

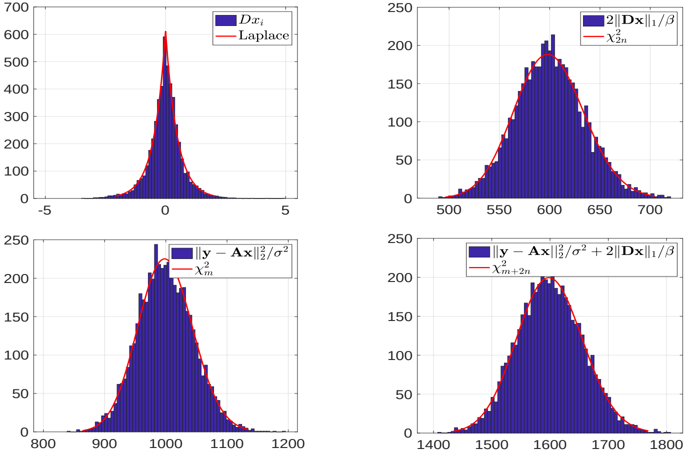
If $z_i \sim \text{Laplace}(\theta, \beta)$ all independent, then

$$\sum_{i=1}^{n} \frac{2|z_i - \theta|}{\beta} \sim \chi^2_{2n}.$$

Since the TV functional is differentially Laplacian

$$\frac{\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2}{\sigma^2} + \frac{2\|\mathbf{D}\mathbf{x}\|_1}{\beta} \sim \chi^2_{m+2n}$$

Histograms illustrating χ^2 Test for TV Functional



χ^2 Test for TV

$$\frac{\|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}_{tv}(\lambda)\|_2^2}{\sigma^2} + \frac{2\|\mathbf{D}\hat{\mathbf{x}}_{tv}(\lambda)\|_1}{\beta} \sim \chi^2_{m-df(\mathbf{A}\hat{\mathbf{x}}_{tv}(\lambda))+df(\mathbf{D}\hat{\mathbf{x}}_{tv}(\lambda))}$$

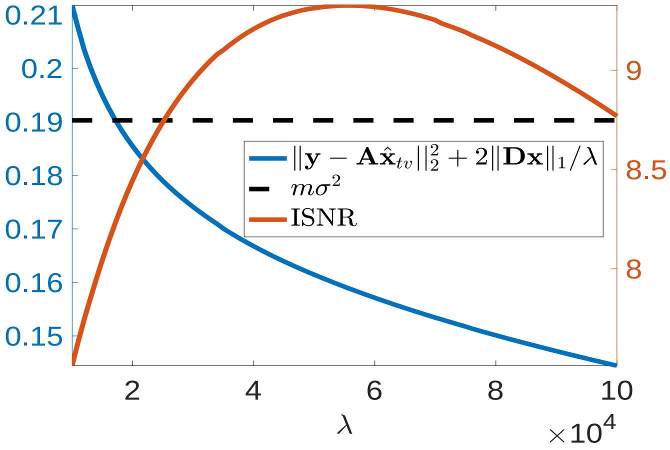
Theorem

*Suppose that $(\mathbf{y} - \mathbf{A}\mathbf{x})_i \sim \mathcal{N}(0, \sigma)$ and $(\mathbf{D}\mathbf{x})_i \sim$ Laplace(θ, β) with $\mathbf{A}$ and $\mathbf{D}$ full rank. Then

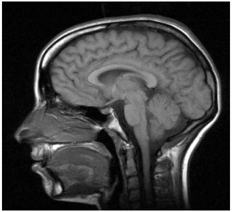
$$\frac{\|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}_{tv}(\lambda)\|_2^2}{\sigma^2} + \frac{2\|\mathbf{D}\hat{\mathbf{x}}_{tv}(\lambda)\|_1}{\beta} \sim \chi^2_m$$

*Mead 2020

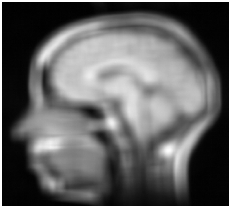
MRI BSNR = 40; χ^2 ISNR = 8.22; Max ISNR = 9.33



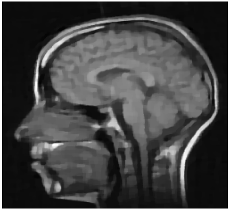
Exact



BSNR $= 40$



χ^2 ISNR $= 8.22$



Maximum ISNR $= 9.33$

Summary and Conclusions

- We have developed a framework for automatic and efficient selection of TV regularization parameters. The approach extends results on residuals and risk estimators, in particular
  - The new measure of risk involves the regularized residual which follows a χ^2 distribution.
  - The degrees of freedom can be estimated from recent results on degrees of freedom for generalized Lasso.
- The proposed TV regularization parameter selection method* requires a data noise estimate and solves the TV problem multiple times during an optimization, rather than guess and checking.

*Mead 2020